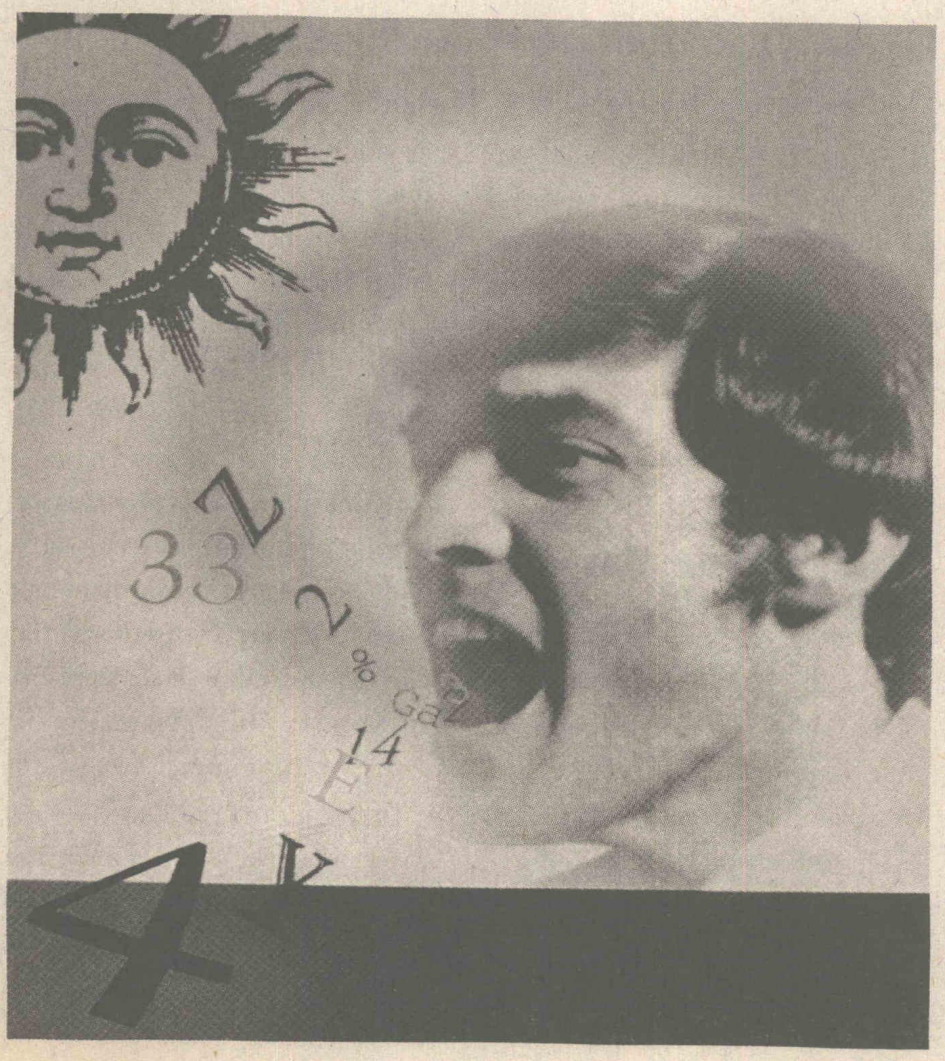
กิตติพงษ์ มุ่งนิรันดร์

การติดต่อ
กับผู้ใช้ด้วยเสียง

ปากว่ามือถึง กับเครื่องคอมพิวเตอร์ ด้วยการใช้เสียงพูดติดต่อสั่งงานเครื่องคอมพิวเตอร์ด้วย VOICE-INTERFACE SYSTEM



ในยุคปัจจุบันที่คอมพิวเตอร์มีความสำคัญมากขึ้นในทุกวงการ แต่การติดต่อกับคอมพิวเตอร์โดยการใช้คีย์บอร์ดหรือเม้าส์ก็ไม่ใช่เรื่องง่าย ๆ สำหรับผู้ใช้โดยทั่วไป ระบบการติดต่อกับผู้ใช้ด้วยเสียงจะให้วิถีทางใหม่สำหรับผู้ใช้ การติดต่อด้วยภาษาพูดเป็นรูปแบบการสื่อสารที่ถือได้ว่าดีที่สุดสำหรับมนุษย์กับคอมพิวเตอร์

เทคโนโลยีด้านการจดจำเสียงพูดจะทำให้อุปสรรคกีดขวางระหว่างคนกับเครื่องจักรหมดไป มันจะทำให้ทุก ๆ คนสามารถใช้คอมพิวเตอร์ได้อย่างง่ายดาย และทำให้ผู้ที่ใช้คอมพิวเตอร์เป็นประจำสามารถใช้ได้อย่างมีประสิทธิภาพมากขึ้น

นอกจากนี้ การใช้เสียงพูดยังมีข้อดีอีกมากมาย การป้อนข้อมูลทำได้อย่างรวดเร็ว เพราะว่าโดยทั่วไป คนเราสามารถพูดได้โดยเฉลี่ยถึง 200 คำต่อนาที แต่การใช้คีย์บอร์ดทำได้อย่างดีก็แค่ 60 คำต่อนาที และในสถานการณ์หรือสถานที่ที่มีอุปสรรคในการใช้เครื่องคอมพิวเตอร์ เช่น การทำงานในที่มืด การทำงานที่มือไม่ว่าง การขับรถยนต์ และอื่น ๆ อีกมาก การสั่งงานด้วยเสียงพูดทำให้อุปสรรคเหล่านี้หมดไปอย่างง่ายดาย

รูปแบบของเทคโนโลยีด้านเสียง

เป็นเวลามากกว่ายี่สิบปี ที่ได้มีการค้นคว้าวิจัยด้านการจดจำเสียงพูด

รูปแบบต่าง ๆ สามารถเห็นได้อย่างชัดเจนมาก ได้แก่

- การขึ้นตรงต่อผู้พูด (Speaker dependence) และการไม่ขึ้นตรงต่อผู้พูด (Speaker independence)

ระบบขึ้นตรงต่อผู้พูดจะต้องได้รับการฝึกเพื่อให้จดจำเสียงพูดเฉพาะบุคคลไป และระบบที่ไม่ขึ้นตรงต่อผู้พูดสามารถจดจำเสียงของใครก็ได้ แต่ความถูกต้องจะน้อยกว่า

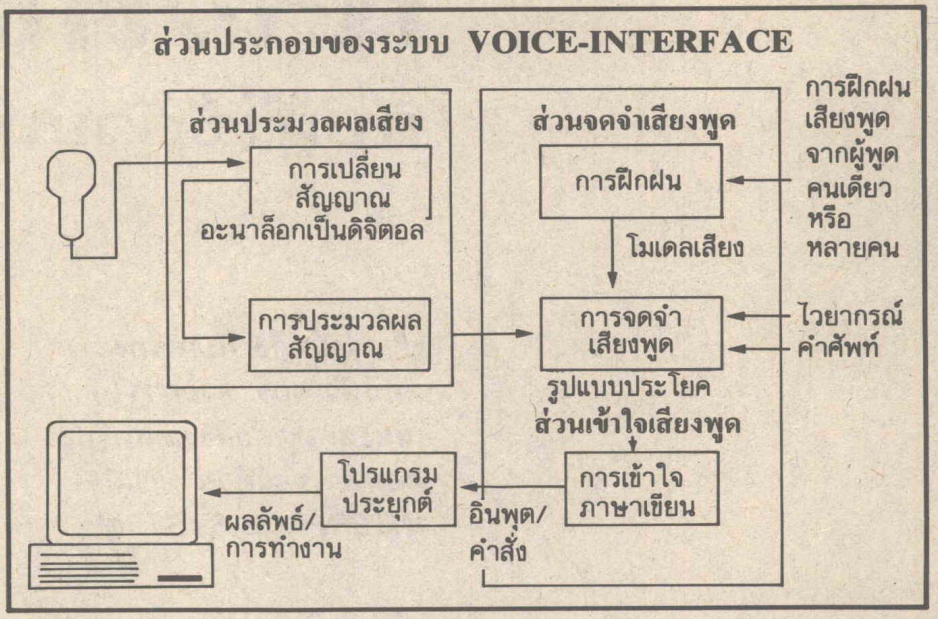
- การพูดแบบเป็นคำ ๆ (Discrete words) และการพูดแบบต่อเนื่อง (Continuous speech)

ระบบพูดแบบเป็นคำ ๆ จะต้องมีหยุดพูดระหว่างคำ ส่วนระบบพูดแบบต่อเนื่องสามารถรับคำพูดได้แบบต่อเนื่องในลักษณะที่เหมือนธรรมชาติมากกว่า แต่มีข้อเสียคือ ระบบจะมีความซับซ้อนและความผิดพลาดมากกว่า

- ขนาดของคำศัพท์และความซับซ้อนของไวยากรณ์

ขนาดของคำศัพท์ที่เครื่องสามารถจดจำได้และไวยากรณ์ที่กำหนดรูปประโยคที่เครื่องสามารถแยกแยะได้ ถ้าระบบที่มีจำนวนคำศัพท์น้อยและรูปแบบประโยคจำกัดจะซับซ้อนน้อยกว่า แต่ถ้าเป็นระบบที่มีจำนวนคำศัพท์มากและรูปแบบประโยคหลาย ๆ แบบจะมีประโยชน์มากกว่า

ระบบติดต่อกับผู้ใช้ด้วยเสียง (voice-interface system) ทั้งที่มีอยู่ในท้องตลาดและที่วิจัยพัฒนากันอยู่นั้นต่างก็มุ่งหวังที่จะให้ระบบที่มีความถูกต้องสูง แต่ทุก ๆ ระบบก็ถูกจำกัดไว้ด้วยคุณสมบัติทั้งสามข้อที่ผมได้พูดถึงไปแล้ว ดังนั้นเวลาที่คุณต้องการใช้งานระบบนี้ คุณจะต้องพิจารณาและเลือกอย่างระมัดระวังเพื่อให้เหมาะสมกับโปรแกรมประยุกต์ที่คุณใช้ด้วย



รูปที่ 1 ส่วนประกอบของระบบ VOICE-INTERFACE

ตามรูปที่ 1 แสดงให้เห็นถึงโครงสร้างโดยทั่วไปของระบบติดต่อกับผู้ใช้ด้วยเสียง ทุก ๆ ระบบประกอบด้วยส่วนสำคัญสามส่วนดังนี้ ส่วนประมวลผลเสียงพูด (speech processing) ส่วนจดจำเสียงพูด (speech recognition) และส่วนเข้าใจเสียงพูด (speech understanding)

ส่วนประมวลผลเสียง

ส่วนประมวลผลเสียงพูดทำหน้าที่เปลี่ยนเสียงพูดไปเป็นสัญญาณดิจิตอลด้วยการใช้ Analog/Digital converter และใช้เทคนิคทางด้านการประมวลผลสัญญาณเพื่อกำจัดข้อมูลที่ไม่มีความสำคัญในเสียงพูดออกไปและจัดเก็บให้อยู่ในรูปแบบที่เหมาะสม

ระบบจะทำหน้าที่รับข้อมูลเสียงผ่านทางไมโครโฟนหรืออุปกรณ์อย่างอื่นเข้ามาเป็นสัญญาณทางไฟฟ้าหรือที่เรียกว่า สัญญาณอะนาล็อก (analog signal) และเปลี่ยนเป็นสัญญาณดิจิตอลโดยการใช้ A/D converter โดยทั่วไปอัตราการแปลงข้อมูลจะทำที่ความถี่ตั้งแต่ 6000 ถึง 20000 ครั้งต่อวินาที โดยข้อมูลที่ได้จะมีขนาด 12 บิตต่อ 1 สัญญาณเพื่อให้ได้เสียงที่มีคุณภาพเพียงพอ ทำให้เราได้เสียงเก็บไว้ในรูปแบบที่เรียกว่า digital waveform

โดยหลักการแล้ว เราสามารถทำการวิเคราะห์และจดจำข้อมูลในรูปแบบนี้ได้ แต่เนื่องจากการพูดเพียงแค่ 5 วินาทีจะทำให้เราได้ข้อมูลเสียงออกมาถึง 100000 หน่วย ซึ่งการจดจำเป็นไปได้ยากมาก ดังนั้นจำเป็นต้องใช้เทคนิคด้านการประมวลผลสัญญาณ เพื่อทำการกำจัดข้อมูลที่ซ้ำซ้อนและไม่มีความสำคัญออกไปเสียก่อน เช่น ใช้เทคนิค Fast Fourier Transforms เป็นต้น ทำให้ข้อมูลเสียงเก็บอยู่ในรูปข้อมูลเวกเตอร์ประมาณ 8 ถึง 20 จำนวนในรูปเลขจำนวน (floating-point number) ต่อ

ข้อมูลเสียง 0.01 วินาที นอกจากนี้ยังมีการใช้เทคนิคอื่น ๆ มากมายเพื่อการเก็บเสียงพูดทำได้อย่างมีประสิทธิภาพ

ส่วนจดจำเสียงพูด

ส่วนจดจำเสียงพูดจะทำการค้นหาและจับคู่เสียงที่พูดเข้ามากับโมเดล หรือฐานข้อมูลเสียงซึ่งได้รับการฝึกและเรียนรู้จากตัวอย่างผู้พูด นอกจากนี้ ยังทำหน้าที่แยกแยะคำศัพท์และหลักไวยากรณ์ที่พูด

โดยปกติ ผู้ใช้จำเป็นต้องทดสอบการพูดเพื่อสร้างโมเดลเสียงขึ้นมา ก่อนใช้งานระบบได้ บางระบบจะทดสอบให้พูดแตกต่างกัน เช่น พูดเป็นคำ วลี หรือพยางค์ บางระบบต้องพูดทุก ๆ คำที่จะใช้งาน เมื่อระบบมีขนาดเล็ก และบางระบบพูดเพียงบางคำเท่านั้น

สำหรับเทคนิคด้านการแยกแยะคำศัพท์ว่า คำแต่ละคำที่เราพูดนี้หมายความถึงอะไร นอกจากวิธีที่เปรียบเทียบคำศัพท์ตัวต่อตัวแล้ว ระบบส่วนใหญ่จะเอาการวิเคราะห์ด้านไวยากรณ์และข้อจำกัดของแต่ละภาษามาแยกแยะ เพื่อกำจัดแนวทางที่เป็นไปไม่ได้ออก เช่น เมื่อเราพูดคำว่า เรา ระบบจะทำการแยกสระออกมาก่อนเป็น เ-า เพื่อนำไปใช้เปรียบเทียบเฉพาะคำในกลุ่มนี้เท่านั้น

ส่วนเข้าใจเสียงพูด

ส่วนเข้าใจเสียงพูดทำหน้าที่เข้าใจและเปลี่ยนข้อความเหล่านี้ให้เป็นคำสั่งที่แน่นอนสำหรับสั่งงานเครื่องคอมพิวเตอร์ ระบบในปัจจุบันใช้งานได้ดีกับระบบขนาดเล็กและจำกัดรูปแบบการพูด แต่ยังไม่ดีพอสำหรับการพูดที่ไม่มีการจำกัด

VOICE INPUT FOR PERSONAL COMPUTERS

Company	Product name	PC type	Card/box/stand-alone	Memory required	Speaker independent/dependent	Discrete-word/continuous speech	Maximum active vocabulary	Phone interface	Text-to-speech capability	Voice playback and storage	DTMF (Touch-Tone) functions	Price	Comments
Articulate Systems	Voice Navigator	Mac Plus	Box	2 MB	SD	DW	200	Yes	No	No	Optional	$1295	Toolkit optional; uses Dragon Systems algorithms; connects to SCSI port.
Cherry Electrical Covox	Voice Scribe 1000 Plus	PC/XT/AT	Card	640K	SD	DW	1000	N/R	Optional	No	No	$3600	Uses Dragon Systems algorithms.
Dragon Systems	Voice Master Key	PC/XT/AT	Card	256K	SD	DW	64	No	Optional	Yes	No	$149.95	Toolkit optional.
	Dragon Dictate-30K	386 (AT)	Card	8 MB	SI	DW	25,000	N/R	No	No	No	$9000	Can also be trained to your voice.
Kurzweil Applied Intelligence	Dragon Writer-1000	286 (AT)	Card	512K	SD	DW	1000	N/R	No	No	No	$3600	Multiple speakers can be trained into one model.
	Voice Report	386 (AT) with 80387 coprocessor	Card	10 MB and 40-MB hard disk	SD	DW	5000+	No	Optional	Optional	Optional	$26,500	Price includes PC; fax option also available
Scott Instruments Speech Systems	SIR Model 20	PC/XT/AT	Card	80K	SI	DW	160	N/R	No	Yes	No	$1495	Toolkit is optional.
	DS200	Sun SPARC-Station	Card	8 MB and 140-MB hard disk	SI	CS	40,000	No	Optional	No.	Yes	$33,900	Toolkit included; acceptable accuracy and speed require syntactic language model constraints on active vocabulary.
Texas Instruments	TI Voice Card	PC/XT/AT	Card	256K and two floppy disk drives	SD	CS	50	Optional	Yes	Yes	Optional	$995	Toolkit is extra; also available for Xenix systems with 386 processor.
The Voice Connection	Introvoice-5	PC/XT/AT	Card	64K	SD	DW	250	No	Optional	No	No	$495	Includes PC with minikeyboard and RS-232C port; full miniature version also available.
Verbex Voice Systems	PTVC-756	Portable/PC	Stand-alone	1 MB	SD	DW	250	No	Yes	No	No	$2995	Includes toolkit.
	Verbex 5000	N/A	Stand-alone	N/A	SD	CS	80	N/R	Yes	No	No	$5600	Includes toolkit.
	Verbex 6000	AT	Card	None	SD	CS	300	N/R	Optional	No	No	$5600	Includes toolkit.
	Verbex 7000	N/A	Stand-alone	N/A	SD	CS	1980	N/R	Yes	No	No	$9600	Includes toolkit.
	TeleRec	XT/AT	Card	64K	SI	DW	16	Yes	No	No*	Yes	$2495	*Voice-prompt playback.
	VR-4	PC/XT/AT	Card**	64K	SI	DW	50	Yes	No	No*	Yes	$3600	*Voice-prompt playback. **Card is piggybacked with Dialog/4x voice card from Dialogic Corp.; other daughtercards also available.
Voice Control Systems	VPC-1000	AT	Card	None	SI	CS	13	Yes	No	No	No	$5500	Toolkit included.
Voice Processing Corp.	Voicebox	PC/XT/AT	1/2 card	80K	SD	DW	500	No	No	No	No	$3500	For evaluation kit.
Voice Recognition Technologies												$395	Includes toolkit.
Votan	Voice Card VPC-2100	PC/XT/AT	Card	256K	SD	CS	125	Yes	No	Yes	Optional	$1750	Multibus card available.

Note : This table includes no claims for the accuracy of the voice interface.
N/R = Not recommended by the vendor.
N/A = Not applicable.

รูปที่ 2 ระบบติดต่อด้วยเสียงบน IBM PC และคอมแพทิเบิล

รูปแบบประโยค

สำหรับการประยุกต์ใช้ในบางงาน เช่น การป้อนข้อมูล (Data entry) การสั่งงานด้วยคำสั่ง เป็นต้น เพียงแต่ต้องการให้ระบบจดจำได้ถูกลำดับของคำที่พูดออกมาเท่านั้น ไม่จำเป็นที่จะต้องเข้าใจถึงรูปประโยคและความหมายมากนัก แต่ในงานบางอย่าง ระบบต้องเข้าใจและสามารถแยกความแตกต่างของแต่ละประโยคและความหมายได้ ตัวอย่าง เช่น ในการถามข้อมูลว่า "จำนวนลูกค้าใหม่ในหน่วยงานมีเท่าไร" จำเป็นจะต้องประมวลผลที่ซับซ้อนพอดู เพื่อให้เข้าใจถึงคำว่า "ใหม่" นอกจากนี้ ระบบไม่เพียงแต่พิจารณาถึงรูปแบบประโยคของประโยคคำถามปัจจุบันเท่านั้น แต่จะต้องต่อเนื่องและสอดคล้องกับคำถามและคำตอบก่อนหน้านี้ด้วย

ในปัจจุบัน การเข้าใจเสียงพูดและภาษาแบบไม่จำกัดรูปแบบประโยคเป็นสิ่งที่ยากมากและยังไม่สามารถหาวิธีการที่เหมาะสมออกมาได้ แต่สำหรับระบบที่มีการจำกัดรูปแบบประโยคสามารถเป็นไปได้

การใช้งานระบบจดจำเสียง

ระบบติดต่อด้วยเสียงไม่ใช่เป็นเรื่องอนาคตอีกต่อไป เพราะได้มีระบบนี้อยู่บน IBM PC/XT และเครื่องคอมแพตทิเบิลแล้ว ดูได้จากรูปที่ 2

การ์ดด้านเสียงนี้มีการแบ่งประเภทออกตามลักษณะของวิธีการ ได้แก่ การขึ้นตรงและไม่ขึ้นตรงต่อเสียงพูด การพูดแบบเป็นคำ ๆ และแบบต่อเนื่อง ขนาดของคำศัพท์ และคุณสมบัติอื่น ๆ ที่เพิ่มเติมเข้ามา ผมได้เคยทดลองใช้การ์ดด้านเสียงนี้มาบ้างแล้ว คือ TI Voice Card ของบริษัท Texas Instruments เป็นอุปกรณ์ที่มีขนาดไม่ใหญ่มากนัก มีการต่อลำโพงและไมโครโฟน การรับเสียงพูดเป็นแบบขึ้นตรงต่อผู้ใช้ กล่าวคือ ต้องมีการพูดคำสั่งทั้งหมดที่จะใช้ไว้ก่อนและเก็บข้อมูลลงไฟล์ไว้ก่อนโดยในภายหลังสามารถนำออกมาใช้ได้อีก จำนวนคำศัพท์เก็บได้น้อย แต่ในโปรแกรมประยุกต์ที่ในเวลาหนึ่ง ๆ มีการใช้เสียงเพียงไม่กี่คำสั่ง เพราะเราสามารถโหลดชุดคำสั่งที่เหมาะสมขึ้นมาแทนที่ได้ ก็นับได้ว่าเป็นอุปกรณ์ที่น่าสนใจมาก

ที่น่าสนใจมาก ๆ ได้แก่ การนำเอาการ์ดเหล่านี้ไปใช้ร่วมกับโปรแกรมประยุกต์ต่าง ๆ ที่มีใช้อยู่ในท้องตลาดอยู่แล้ว เช่น โปรแกรมสเปรดชีต โปรแกรมฐานข้อมูล เป็นต้น โดยแทนที่จะต้องคีย์ข้อความคำสั่งลงไป ก็เพียงแต่พูดคำสั่งและระบบจดจำเสียงจะทำการเปลี่ยนไปเป็นชุดคำสั่งหรือตัวหนังสือจำนวนมากมายได้ด้วยคำพูดเพียงคำเดียว โปรแกรมประยุกต์เหล่านี้ก็จะได้รับคีย์ที่ส่งเข้ามาจากระบบเหมือนกับว่าได้รับมาจากคีย์บอร์ด

การประยุกต์ใช้ด้านอื่น ๆ อีกมาก เช่น นำไปใช้ในโรงพยาบาล เพื่อให้คนป่วยที่ไม่สามารถเคลื่อนไหวได้สั่งงานด้วยเสียงพูดแทน คนพิการสามารถใช้งานคอมพิวเตอร์ได้โดยใช้คำพูดแทนการคีย์ข้อความ นำไปใช้ในงานวิจัยที่ต้องการใช้เครื่องคอมพิวเตอร์ในขณะที่มือไม่ว่างหรือทำงานอยู่

การประยุกต์ใช้ด้านอื่น ๆ

การติดต่อด้วยเสียงไม่ใช่มีประโยชน์เพียงกับเครื่องคอมพิวเตอร์เท่านั้น การประยุกต์ที่น่าสนใจในรูปแบบอื่น ๆ คือ การนำเอาเทคโนโลยีด้านเสียงไปใช้ร่วมกับโทรศัพท์ ระบบมัลติมีเดียและการส่งข้อความเป็นเสียง (voice mail) ซึ่งเป็นการประยุกต์ของระบบ e-mail (electronic mail)

ในขณะนี้ เครื่องบางระบบเช่น NeXT-Computer, SUN-Workstation และระบบที่ใช้ X Window มีระบบ mail ที่สามารถส่งข้อความเป็นเสียงพูดรวมไปกับข้อความธรรมดาได้ โดยในเครื่องที่มีระบบนี้เหมือนกันสามารถที่จะฟังเสียงนี้ได้

อนาคตของระบบ

ในอนาคต ท่านจะได้พบว่า ระบบคอมพิวเตอร์จะมีความสามารถทางด้านเสียงมากขึ้นและอาจจะกลายเป็นอุปกรณ์มาตรฐานรวมเข้าไปในระบบด้วย โปรแกรมประยุกต์ในอนาคตจะรวมเอาความสามารถนี้เข้าไปเป็นส่วนหนึ่ง เช่น ฐานข้อมูลที่เก็บเสียง ระบบ electronic mail โปรแกรมประมวลผลข้อมูล เป็นต้น แต่ในขณะนี้ ผลิตภัณฑ์ยังต้องมีการฝึกและทดสอบเสียงพูดเพื่อให้ระบบปรับตัวเข้ากับผู้ใช้แต่ละคนก่อนนำไปใช้ได้จริง แต่ในอนาคตอันใกล้ก็คงจะมีผลิตภัณฑ์ที่มีประสิทธิภาพสูงขึ้นอย่างแน่นอน